

Red flags in data: Learning from failed data reuse experiences

Ayoung Yoon

Indiana University Purdue University Indianapolis
(IUPUI)
535 W. Michigan St., IT 563
ayyoon@iupui.edu

ABSTRACT

This study examined the data reusers' failed or unsuccessful experience to understand what constituted reusers' failure. Learning from failed experiences is necessary to understand why the failure occurred and to prevent the failure or convert the failure to success. This study offers an alternative view on data reuse practices and provides insights for facilitating data reuse processes by eliminating core components of failure. From the interviews with 23 quantitative social science data reusers who had failed data reuse experiences, the study findings suggest: (a) ease of reuse, particularly the issue of access and interoperability, is the important initial condition for a successful data reuse experience; (b) understanding data through documentation may be less of an issue, at least for experienced researchers to make their data reuse unsuccessful, although the process can still be challenging; and (c) the major component of failed experience is the lack of support in reusing data, which emphasizes the need to develop a support system for data reusers.

Keywords

Data reuse, data practice.

INTRODUCTION

Data sharing and reuse have become increasing concerns in many scientific disciplines. While there have been growing interests in data sharing among researchers because of the acknowledged benefits of data sharing and mandates by major funding agencies' requirements, it is important to note that data sharing does not automatically lead to data reuse. Despite that the underlying intention of data sharing is data reuse, Peer, Green, and Stephenson (2014) argued

that simply sharing data is not sufficient for future reuse. Many researchers pointed out the complex nature of data reuse processes (Carlson & Anderson, 2007; Zimmerman, 2008), and highlighted the need to make the data as an interpretable and in as usable form as possible so it can be reused (Peer et al., 2014). Data also need to be reasonably or properly processed, shared, and preserved to be in the condition of being interpretable and usable. However, the terms "reasonably" or "properly" are not easily defined and are contested concepts (Carlson & Anderson, 2007). Further, what these terms mean to data reusers can be different depending on specific data practices, types of data interacted with, disciplinary culture, and institutional environments. Examining data reusers' experiences and practices is important in this regard, as it informs how to prepare and manage data for future reuse from various reusers' perspectives and needs.

As a response, an increasing number of studies has investigated researchers' data reuse practices in different disciplines, explored the different nature of data reuse in different contexts, and reported various challenges associated with reuse experiences (e.g., Carlson & Anderson, 2007; Faniel & Jacobsen, 2010; Niu, 2009; Palmer, Weber, & Cragin, 2011; Yoon, 2014, 2016a; Zimmerman, 2007). While each component of these previous studies has enhanced the understanding on the full data reuse lifecycle, most studies addressed the challenges and difficulties during the process of reuse. Few studies focused on data reusers' failed or unsuccessful data reuse experiences, which resulted in non-use of existing data. Learning from failure is necessary to understand why the failure occurred and to draw lessons to prevent problems that caused the failure as a basis for planning future success. The goal of this study is, thus, to investigate the components that cause researchers' failed or unsuccessful data reuse experiences.

This study focuses on quantitative social science data reuse that has a long history of reuse culture. Failed data reuse experience is defined as the experiences using existing data that do not lead to any intended research outcome (e.g., journal publications, conference presentations, or research reports) and experiences in which researchers had to stop

{This is the space reserved for copyright notices.}

ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.

[Author Retains Copyright. Insert personal or institutional copyright notice here.]

This is the author's manuscript of the article published in final edited form as:

Yoon, A. (2016). Red flags in data: Learning from failed data reuse experiences. *Proceedings of the Association for Information Science and Technology*, 53(1), 1–6. <https://doi.org/10.1002/pr2.2016.14505301126>

using data for any reason. To be considered as a failed experience, researchers had already conducted an initial screening on data for the relevancy and fitness to reuse, and intended to reuse the data for their research projects. Simply changing their mind or finding a better dataset was not counted as a failed experience in this study. From the interviews with 23 quantitative social science data reusers who had failed data reuse experiences, this study will contribute to the existing data reuse literature by adding a new perspective and revealing another dimension of data reuse practices.

LITERATURE REVIEW

Although both Rolland and Lee (2013) and Palmer et al. (2011) argued that there has been little study on actual data reuse practices and called for more research, the number of empirical research studies on data reuse practices has been growing in recent years. Previous studies have investigated data reusers' perceptions, experiences, and attitudes toward data reuse by focusing on specific disciplines or scientific communities, such as engineering (Howard, Darlington, Ball, Culley, & McMahon, 2010), earthquake engineering (Faniel & Jacobsen, 2010), astronomy (Sands, Borgman, Wynholds, & Traweek, 2012), ecology (Zimmerman, 2008), cancer epidemiology (Rolland & Lee, 2013), archaeology (Faniel, Kansa, Whitcher Kansa, Barrera-Gomez, & Yakel, 2013), general social sciences (Faniel, Kriesberg, & Yakel, 2015; Yoon, 2014, 2016a, 2016b), or by comparing different disciplines or scientific groups (Birnholtz & Bietz, 2003; Borgman, Wallis, & Mayernik, 2012; Carlson & Anderson, 2007; Kriesberg, Frank, Faniel, & Yakel, 2013). These studies revealed that data reuse practices are varied in disciplinary context or the types of data that researchers interact with, and suggested that a one-size-fits-all solution to understand data reuse practices may not be possible.

Despite these variances and differences, these previous studies presented common natures of data reuse as well as fundamental challenges that appeared from data reuse practices. First, data reuse is a complex and contextualized process and reusers work hard to understand the data they try to reuse (Zimmerman, 2008); thus, data reuse is an iterative and ongoing process (Rolland & Lee, 2013). Data are also highly social, with embedded values and practices of researchers who are involved in and engaged with data creation (Zimmerman, 2008), which reveals the contextual nature of data. Carlson and Anderson (2007) argued raw data were "never self-explanatory or self-legitimizing" (p. 647), which demonstrated the need to document the degree to which and how the data were constructed. However, documenting contextual information about the data is known to be challenging for two reasons. First, it is simply impossible for researchers to document every decision they made on data as well as tacit knowledge they used in data creation (Niu, 2009; Rolland & Lee, 2013; Zimmerman, 2008). Second, it is difficult to know what would be considered to be enough context information to fulfill data

reusers' different needs and expectations (Birnholtz & Bietz, 2003; Carlson & Anderson, 2007). Birnholtz and Bietz (2003) pointed out the innate challenges of data reuse as knowledge reuse by stating: "Knowledge transfer in this instance is not simply a matter of sharing a set of instructions, but is a highly social process of learning practices that are not easily documented" (p. 341).

While previous studies on data reuse contribute to the better understanding of data reuse, which is a foundation to improve the current practices through necessary human, technical, and systematic supports, none of these studies has investigated the topic from the perspectives of researchers who have not been successful in data reuse. This study fills the gaps in existing literature by addressing data reusers' failed or unsuccessful experiences. Exploring failed experiences will expose unseen or un-emphasized aspects of successful data reuse experiences in which researchers had to deal with difficulties. By understanding what causes researchers' failed experience, this study will contribute by providing some insights for researchers to lower the barriers to entry to data reuse.

METHODS

The primary data were collected through in-depth, semi-structured interviews with 23 quantitative social science data reusers from social work and public health who have experienced unsuccessful data reuse. Failed or unsuccessful data reuse experience is defined as the experiences reusing existing data that do not lead to any intended research outcome and experiences in which researchers have to stop using data for any reason. To find researchers with failed data reuse experiences, The PI first identified researchers who had success in data reuse, meaning those who were able to publish any type of research outcome by reusing existing data, and then screened them based on failed experiences. This sampling strategy is, first, for the convenience of the sample. Finding researchers who had a failed data reuse experience is challenging because these experiences are usually not documented and thus are hard to track. Second, by identifying researchers who have not had just a failure but also a success in data reuse, this study tries to minimize individual variances in failed experiences from the lack of experiences and exclude a failure from first-time trials. Individual researchers' skills, knowledge, background, and level of experiences and training with data can be varied and an important factor in failed data reuse. However, by addressing researchers with at least some experience in data reuse, this study will focus more on conditions of and issues with data available to researchers.

Study participants were identified from the major social science databases (e.g., EBSCOHost, SAGE Journals, ProQuest Social Science, and ERIC). Potential participants were searched using several keywords, such as "secondary data" or "secondary analysis" in a full text search. The searches were limited to journal publications, conference proceedings, and reports published in the United States for

the convenience of conducting interviews. Authors' information was collected from the database information as well as an additional Google search, and the invitation to the study was sent to the first author or corresponding authors of the article.

Phone interviews with 23 data reusers were conducted from May to September 2014. Average length of the interview was an hour. During the interviews, nondirective open questions were asked to the participants, focusing on their failed reuse experiences and any factors that influenced the failure. Participants were also encouraged to include their successful experiences as a comparison to their failed experiences. All interview data were audio recorded and fully transcribed. The data were openly coded and emerging themes from the analysis were captured using NVivo, a qualitative analysis software.

Study participants

The participants of this study were mostly researchers with PhDs in different positions of assistant, associated, full professors, or research scientists, except for three PhD students who were close to finish their degrees. The participants were mixed in their genders (6 male participants; 17 female participants) and ages (from 20 to 70). All participants had more than one successful data reuse experience, meaning they had more than one publication result that came out of the project that reused existing data. The minimum number of datasets the participants interacted with was two (one successful and one unsuccessful experience). The datasets the participants had reused were a mix of publically available data and data they received from personal exchange.

RESULTS

Even though the study participants had at least one successful experience of reusing data, they reported unsuccessful experiences as well, including the case in which they had to stop reusing data after they found the data and decided to reuse. The point that each participant stopped reusing data during their research cycle was varied; some stopped at the very early stage of reuse and others had to stop after spending a significant time on the data, which made the participants unhappy as "it's a lot of time wasted" (I11).

Wrong or incomplete descriptive information

A few participants said that they had to not reuse data after initial exploration of the data, as they found that the actual data (variables, measurements, etc.) did not exactly match with the descriptive information about the data they found during their search process. This was not any problem within the actual data or with data reusers' ability to retrieve information relevant to the data, but rather it was wrong or incomplete descriptive information about the data. I05 said, "I usually start considering data that I already know what's in there..., but the only times that [not using data] happened that when I found the data didn't have the

variables or sample size that I thought the dataset had." I02 and I13 reported a similar experience; later, they both found the data were older than they thought and needed. Both considered that was a simple mistake of whoever prepared the description of data, but sometimes "looking at the data [themselves] was kind of challenging as well, cause you don't know unless you really dig into [them]" (I13).

All of the datasets reported as having an issue with description were publically available data, retrieved from either nonprofit data repositories or research institutions' own repositories. It does not indicate that the description about data from individual researchers was perfect or complete, but because only publically available data provide a public view on the description about the data that can be searchable through different catalog systems and metadata. Participants tended to check the interested variables or measures through in-person communication before they received the data directly from other researchers, which caused fewer issues in mismatch between data description and actual data.

Difficult access to data

The issue of access was reported as a reason to not reuse data anymore by the most participants. The participants acknowledged that the access issue may not be a reason for unsuccessful data reuse for all researchers depending on their needs and willingness to invest extra effort. I22 said "it becomes that tradeoff."

I22: It's one of those cost-benefit analyses kinds of things, right? ... There's a good number of hoops to jump through to get access to [the data]. ... [Because] sometimes, people have very short timetables they're working on for whatever reason, it might not be worth it to invest all that time and energy to gaining access and getting set up with the data.

Several participants gave up reusing data due to the difficulties in accessing data. I08 considered the easy access as the most important condition of data reuse, and said "as long as [the data] are easily accessible, I felt like I could handle [other issues that may come up], ...but [the data] were harder to access." I10 reported an experience of noticing the route to the data access that she was used before had been changed.

I10: I don't enjoy using [the data] as much I used to because the [data] webpages and the way they give you the data is much more complex than it used to be. ... I've actually stopped including them in [my project]. Well, the dataset itself is probably similar, but the way you get to it is just not easy anymore. It's so frustrating that it's just really hard to use.

Data formats and software

Several participants reported difficulties with unfamiliar data format, software, or special analytic programs, which eventually prevented them from reusing the data. I01 struggled to download some data because of the format: "There're also data sources that are incredibly labor intensive to get the data downloaded and functioning. [These data] have different data formats.... [I]t just becomes a huge headache." I08 was also unable to open data because "there's a kind of, some process involved." She was not sure of the format of the data or if she needed a special program to open and run the data. She needed "some assistance from a statistician or other experts," but she ended up abandoning the data, thinking, "It's just too much," without any institutional support or data services available to her. I18 had tried to use some data from Europe: "You can't just use regular SPSS [Statistical Package for the Social Sciences] for that data. You have to use programs that are set up for [it]" (I18). She had to use the analytic program designed for the data format but was not sure if the new software program was worth her time and money. Participants would use the data if the data were "at least transfer[able] pretty easily" (I05), and they were "probably a little reluctant to use [some data] because you do have to use [special] statistical software..." (I18), depending on the research support level at their school or institution.

Navigating data through documentation

One participant had a difficulty navigating variables within the dataset. I13 recalled the memory of navigating "a large, vast dataset," which was recommended by her advisor. She found "the documentation on the variables is not as easy to navigate," and she eventually "gave up." Documentation appeared to be a main issue in navigating the variables included in the dataset. There is a possibility that other researchers were able to use those data despite the difficulties, given the fact that I13's advisor once recommended data to review, and it is also possible there are discrepancies in individual researchers' level of ability to navigate complex data.

Questionable values in data

While the several issues reported previously were found relatively early during the participants' investigations on the data, participants found other issues much later during the process of data reuse while they were closely working with the data and running analyses. Three participants said that they found either too many unexpected missing values in the data or problems with samples (missing large portion of the sample), and also explained why these were problematic to continue to reuse the data.

I11 said, "I didn't have a lot of faith in this particular dataset [because] for the [variables] that I was particularly interested, there was just too much missing data." I03 had a similar experience, but she found the interaction with original investigators were more problematic: "There

[were] a couple of problems. One, there was missing data in the dataset, and the researcher said, 'Well just go through it in the best of your abilities solving the missing data.' So that was a red flag right away." How to handle the missing data was more important for I03, but a good explanation was not provided by the original investigators, which prevented I03's further work with the data.

I08 was more suspicious about the data, particularly for the representativeness of a certain sample population included in the data. She guessed "[The original investigators] did not capture all like [ethnic group A] people." The dataset was well known in the field and consisted of only data available to scholars, according to I08. She did not "think all other scholars are aware of [the problem with samples though] because [you'd know] if you are studying a specific party. So I was really focusing on the same statistics [as well as populations]; that's why I got to know."

Problem with original data analysis, manipulation and cleaning

Four participants shared their experience of stopping data reuse due to the issue with the process of original data manipulation, cleaning, and analysis. I16 reported the issue with data coding and said, "as I looked at that more and how they coded things, I thought it was a bit off." After digging into the several codes, I16 found "They're saying this group is both [A identity] and [B identity]. Well, that's kind of a contradiction. ... So at that point I ended up kind of rejecting [original] codes." I07 was confused whether the problem she had was from data entry or data cleaning, as she found "[The] values that just don't make sense [when] I was doing the analysis." Although she was willing to clean up the data by herself, she recalled that she gave up at certain points as she frequently ran into same problems. Similarly, I17 found the "[the] dataset had been singly imputed, that's not a very accurate way of imputing data. So I didn't use that dataset. And so I've had people tell me about datasets that have that."

Interestingly, perhaps because they already spent some time to understand the data and were at the point of analysis when they found these problems, they usually tried to resolve the problems either by contacting the principle investigators or third party organizations/people (e.g., data repositories or institutional data services). Unfortunately, this was usually unsuccessful. I07 said:

I07: I started making lists of questions and issues in the data, and sending it to [the principle investigators] for clarification. If they can clean it up, good; if not, then it throws a red flag on the whole. ... [The principle investigators] can't respond and figure out what was wrong with it. ... It just says that they had not looked at their data closely enough.

DISCUSSION

The study examined researchers' failed data reuse experiences to offer an alternative view on data reuse practices. The study findings suggest: (a) ease of reuse, particularly the issue of access and interoperability, is the important initial condition for a successful data reuse experience; (b) understanding data through documentation may be less of an issue, at least for experienced researchers to make their data reuse unsuccessful, although the process can still be challenging; and (c) the major component of failed experience is the lack of support in reusing data, which emphasizes the need to develop a support system for data reusers.

Ease of use: A critical starting point for successful data reuse

This study results show that ease of data reuse can be a significant barrier in successful data reuse. Previous studies already pointed out the issues relevant to ease of use in data reuse (Faniel & Jacobsen, 2011; Faniel et al., 2015). Similarly, the easy access to data seems to be the most critical starting point of successful data reuse. The fact that I10 stopped reusing a same dataset after the route to data had been changed demonstrates the significance of easy access. As experienced researchers in data reuse, the study participants also had some level of confidence in handling other issues that may come up once they can access to data, which implied they considered access as the major condition to reuse data. Considering the participants had experiences in data reuse, access can possibly be a bigger barrier for new or novice researchers with little or no experience in data reuse, and thus, is important to provide easy access to data to promote data reuse to inexperienced researchers.

The issue of interoperability in data formats and software should also be addressed for successful data reuse. While study participants were familiar with some common types of quantitative data formats and analytic programs, a few found special formats of data that required special software or analytic programs. These special case data tended to be a problem when data were from another country. Because data will and can be internationally exchanged, interoperability of data across different countries can be a real issue unless it is addressed when developing data reuse cultures.

Documentation: Incidental to "failed" data reuse

A number of past studies have argued that data reuse requires an in-depth understanding of data's context of creation (Carlson & Anderson, 2007; Faniel, Kriesberg, & Yakel, 2012; Zimmerman, 2008), and documentation is known as one good way of transferring contextual information and is significant in satisfactory data reuse experiences for researchers (Faniel et al., 2015; Niu, 2009). Interestingly, only one participant reported an issue with documentation and stopped reusing data for that reason. A few study participants also reported anecdotes about the

issue with documentation, but it appeared as incidental to failed data reuse in this study. This finding does not indicate that understanding data through documentation is not an issue or challenge in data reuse.

Several explanations are possible. As researchers with experiences in data reuse, the participants were confident working with data in general and felt that they "could handle other issues [that may come up]" (I08) during the process of data reuse, such as an issue with documentation. Some participants also talked about other (successful) reuse experiences in which they were able to resolve the issues in documentation through several trials and using their tacit knowledge, though the process was not always smooth. Conditions of documentation can be another factor of this result, although this study could not examine the actual quality of documentation that the participants interacted with. However, the condition of documentation might be a real barrier for inexperienced researchers in data reuse, as novice data reusers made efforts to understand and work with data through different methods (Faniel et al., 2012).

Layers of support: Converting failure to success

A critical component of failed data reuse experiences mentioned by multiple participants was the lack of support, either from institutions, communities, or individuals (mostly referring data producers), during the process of data reuse. Researchers would not reuse data if the problem was innate in the data, such as the issue with validity or reliability. However, if the problems were with technical aspects, data cleaning, and manipulation, participants at least tried to resolve the problems they faced and sought out external help. When I08 and I18 struggled with unfamiliar types of data and analytic programs, they complained about the lack of institutional support for these formats, as well as a lack of data services they could consult with. I03 and I07 were also willing to work on the issues in the data by contacting data producers with a list of questions, but the unsuccessful trials prevented further progress in data reuse. This finding suggests that there is room for some failed data reuse experiences to be converted to successful experiences, if proper supports are provided from individuals, communities, and institutions.

CONCLUSION

This study examined the data reusers' failed experience to understand why it constituted reusers' failure. As noted earlier, learning from failed experiences is necessary to prevent these failures and draw implications to convert the failure to success. The results of this study would be useful to facilitate data reuse processes by eliminating core components of failure.

The current study, however, also has a limitation. As discussed earlier, the failure components might be minimal in this study context, as this study only addressed experienced researchers in data reuse. Researchers with no or less experiences in data reuse may have other reasons for

their failed data reuse experiences. Studying novice data reusers' failed experiences and perspectives would be necessary to understand the full components of failure and to lower the barriers to the entry to data reuse. Comparison between novice and expert data reusers should also be completed, as it will suggest skill sets and knowledge that are essential in data reuse, which will be useful for instructional design and education for novice or new data reusers.

REFERENCES

- Birnholtz, J. P., & Bietz, M. (2003). Data at work: Supporting sharing in science and engineering. In *ACM Conference on Supporting Group Work* (pp. 339–348). Sanibel Island, FL.
- Borgman, C. L., Wallis, J. C., & Mayernik, M. S. (2012). Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work (CSCW)*, 21(6), 485–523. <http://doi.org/10.1007/s10606-012-9169-z>
- Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, 12(2). Retrieved from <http://jcmc.indiana.edu/vol12/issue2/carlson.html>
- Faniel, I., Kansa, E., Whitcher Kansa, S., Barrera-Gomez, J., & Yakel, E. (2013). The challenges of digging data: A Study of context in archaeological data reuse. In *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 295–304). New York, NY, USA: ACM. <http://doi.org/10.1145/2467696.2467712>
- Faniel, I., Kriesberg, A., & Yakel, E. (2015). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, (Early view). <http://doi.org/10.1002/asi.23480>
- Faniel, I., Kriesberg, A., & Yakel, E. (2012). Data reuse and sensemaking among novice social scientists. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–10. doi: 10.1002/meet.14504901068
- Faniel, I., & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW)*, 19(3–4), 355–375. <http://doi.org/10.1007/s10606-010-9117-8>
- Howard, T., Darlington, M., Ball, A., Culley, S., & McMahon, C. (2010). Opportunities for and barriers to engineering research data re-use. other. *University of Bath*. (ERIM Project Document; ERIM Project Document)
- Hinds, P. S., Vogel, R. J., & Clarke-Steffen, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, 7(3), 408–424. doi:10.1177/104973239700700306
- Gleit, C., & Graham, B. (1989). Secondary data analysis: a valuable resource. *Nursing Research*, 38(6), 380–381. doi:10.1097/00006199-198911000-00018
- Kriesberg, A., Frank, A., Faniel, I. M., & Yakel, E. (2013). The role of data reuse in the apprenticeship process. In *ASIS&T 2013 Annual Meeting Proceedings*.
- Niu, J. (2009). Overcoming inadequate documentation. In *Proceedings of the Annual Meeting of the American Society for Information Science & Technology (ASIS&T)*. Vancouver, British Columbia, Canada.
- Palmer, C. L., Weber, N. M., & Cragin, M. H. (2011). The analytic potential of scientific data: Understanding re-use value. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–10. <http://doi.org/10.1002/meet.2011.14504801174>
- Peer, L., Green, A., & Stephenson, E. (2014). Committing to data quality review. *International Journal of Digital Curation*, 9(1), 263–291. <http://doi.org/10.2218/ijdc.v9i1.317>
- Rolland, B., & Lee, C. P. (2013). Beyond trust and reliability: reusing data in collaborative cancer epidemiology research. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 435–444). New York, NY, USA: ACM. <http://doi.org/10.1145/2441776.2441826>
- Sands, A., Borgman, C. L., Wynholds, L., & Traweek, S. (2012). Follow the data: How astronomers use and reuse data. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1–3. <http://doi.org/10.1002/meet.14504901341>
- Yoon, A. (2014). “Making a square fit into a circle”: Researchers' experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1–4. <http://doi.org/10.1002/meet.2014.14505101140>
- Yoon, A. (2016a). Data reusers' trust development. *Journal of the Association for Information Science and Technology (JASIST)*. doi: 10.1002/asi.23730
- Yoon, A. (2016b). Visible evidence of invisible quality dimensions and the role of data management. In *iConference 2016 Proceedings*. Philadelphia, PA. <http://doi.org/10.9776/16123>
- Zimmerman, A. (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, 7(1-2), 5–16. <http://doi.org/10.1007/s00799-007-0015-8>
- Zimmerman, A. (2008). New knowledge from old data: The Role of standards in the sharing and reuse of ecological data. *Science, Technology & Human Values*, 33(5), 631–652. <http://doi.org/10.1177/0162243907306704>